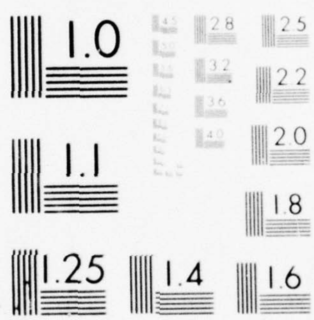
AD-A031 117 CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF STATISTICS F/G 12/2
BAYESIAN ANALYSIS OF THE LINEAR MODEL SUBJECT TO LINEAR INEQUAL--ETC(U)
AUG 76 W W DAVIS N00014-76-C-0930
UNCLASSIFIED TR-118 NL

1 OF 1
AD A031117

END
DATE
FILMED
11-76



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

1

DEPARTMENT

OF

STATISTICS

Carnegie-Mellon University

PITTSBURGH, PENNSYLVANIA 15213

BAYESIAN ANALYSIS OF THE LINEAR MODEL
SUBJECT TO LINEAR INEQUALITY CONSTRAINTS

by

William W. Davis

Technical Report No. 118
ONR Report No. 4
August, 1976

Department of Statistics
Carnegie-Mellon University
Pittsburgh, Pa. 15213

See 1473

Research partially supported by the Office of Naval Research Contract No. N00014-76-C-0930.

1. Introduction

Previous work on linear regression models where the parameter space is restricted by linear inequalities has concentrated on inference from the sampling point of view. For this model the least squares estimate is a solution to a quadratic programming problem [10]. This estimate has a mixed type sampling distribution (i.e., partly continuous and partly discrete) that is difficult to handle analytically, even assuming normal errors. The standard test for significance of coefficients based on the Student-t distribution can be very misleading [12]. Even moments of the estimate are difficult to derive [10,16].

Although much work has been done on Bayesian analysis of regression models, nothing has appeared when the parameter space is restricted. This paper gives an analysis using a natural conjugate prior of the mixed type. Emphasis is placed on determining posterior probabilities that constraints are binding and on determining posterior distributions of the parameters. An analysis is also given for a vague prior of the mixed type.

The basic model is presented in section 2 and examples are given in section 3. The likelihood, the prior, and the posterior are discussed in sections 4, 5, and 6 respectively. Section 7 applies the theory to analyze temperatures of a chemical reaction.

2. The Basic Model

The observations are assumed to follow the standard linear model

$$y = X\beta + e$$

where $y' = (y_1, \ldots, y_n)$, $\beta' = (\beta_1, \ldots, \beta_p)$, $X = (x_1 \ldots x_p)$, and $e \sim N(0, \tau^{-1}I)$. The parameter space

$$\Omega = \{\beta : C\beta \geq f\} \tag{2.2}$$

is assumed to be nonempty, where C is an $r \times p$ matrix.

It will prove convenient in the sequel to separate the unrestricted from the restricted parameters. Thus, we assume that the last $p - p_0$ columns of C are all zeros so that C can be written

$$C = (C_0 \; 0)$$

where 0 is an $r \times (p - p_0)$ dimensional matrix of zeros and $C_0' = (d_1 \ldots d_r)$. If the parameter β is partitioned into restricted and unrestricted parameters, $\beta' = (\beta_1', \beta_2')$ with $\beta_1' = (\beta_1, \ldots, \beta_{p_0})$ and X is similarly partitioned into $(X_1 X_2)$ then the model (2.1) can be written $y = X_1\beta_1 + X_2\beta_2 + e$ with $\Omega = \{\beta : C_0\beta_1 \geq f\}$. The likelihood for the model can be written

$$\ell(\beta, \tau | y) \propto \tau^{n/2} \exp\{-\tau (y - X\beta)'(y - X\beta)/2\} \; I(\Omega) \tag{2.3}$$

where I is the indicator function.

We assume throughout that $p_0 > 0$ since otherwise there are no restricted variables, and the standard Bayesian analysis applies [e.g. 4, section 11.10].

The constraint i $(1 \leq i \leq r)$ is said to be binding if $d_i'\beta_1 = f_i$ where $f' = (f_1, \ldots, f_r)$. In certain applications it is of interest to determine which constraints are binding. Since a Bayesian approach to the problem is adopted here, we compute the posterior probabilities of the various possible sets of binding constraints.

Throughout this paper we will assume that C_0 is of full rank, that $s = \{i_1,\ldots,i_k\}$ is a set of non-binding constraints, and that $\bar{s} = \{i_{k+1},\ldots,i_r\}$ is a set of binding constraints.

We can write the binding constraint equations as

$$C_{\bar{s}}\beta_1 = f_{\bar{s}} \tag{2.4}$$

with

$$C'_{\bar{s}} = (d_{i_{k+1}},\ldots,d_{i_r}) \quad \text{and} \quad f'_{\bar{s}} = (f_{i_{k+1}},\ldots,f_{i_r}).$$

If $r-k \geq 1$, the set of β satisfying (2.4) is a hyperplane of dimension smaller than p. Thus, if the posterior distribution on β is absolutely continuous with respect to p dimensional Lebesgue measure, the posterior probability of the constraint $\bar{s}$ being binding is 0. To alleviate this problem, the prior distribution will be chosen of the mixed type with positive probability on these singular subsets. Dickey [5] gives a bibliography of previous uses of priors of the mixed type.

3. Examples

3.1 Time Series Modelling

The autoregressive process of order p (AR(p))

$$y_t = \sum_1^p \beta_i y_{t-i} + \beta_{p+1} + e_t \qquad t = 1,\ldots,n$$

with $y_0,\ldots,y_{1-p}$ considered as non-stochastic and e_t iid $\sim \mathcal{N}(0,\tau^{-1})$ can be written in the form (2.1) with $\beta' = (\beta_1,\ldots,\beta_{p+1})$ and

$$X = \begin{pmatrix} y_0 & y_{-1} & \cdots & y_{1-p} & 1 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ y_{n-1} & y_{n-2} & \cdots & y_{n-p} & 1 \end{pmatrix} .$$

We will consider the case $p = 2$ (i.e., AR(2)).

The values of (β_1, β_2) determine if the time series is explosive. In fact, the series is explosive if $\beta \notin \Omega$ where Ω is given by (2.2) with $C = \begin{pmatrix} -1 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ and $f' = (-1,-1,-1)$. If one is willing to assume a priori that the time series is not explosive, the assumptions of section 2 are satisfied. Even though the matrix X is stochastic, the likelihood can be written in the form (2.3). Since the theory of sections 4-6 is still valid for stochastic X matrices, the non-explosive AR(2) is an example of the model introduced in section 2.

Zellner [15, section 7.3] computes numerically the posterior probability of the series being explosive assuming $\beta_3 = 0$ and using the vague prior

$$p(\beta_1, \beta_2, \tau) \propto \tau^{-1} .$$

He notes that a similar analysis could be performed using an informative prior, such as the conjugate normal/gamma. Using any prior for (β_1, β_2) that is absolutely continuous with respect to two dimensional Lebesque measure, the posterior probability of any of the constraints being binding is 0. This seems unfortunate since Box and Jenkins [2] demonstrate empirically that models with binding constraints are useful in modelling and forecasting both economic and physical systems. In the present problem $\bar{s} = \{1\}$ corresponds to a difference of order 1 and $\bar{s} = \{1,3\}$ corresponds to a difference of order 2, where the K^{th} difference z_t of the series y_t is given by $z_t = (1-B)^k y_t$ and $B^j y_t = y_{t-j}$. For example, $\bar{s} = \{1\}$ implies $\beta_1 + \beta_2 = 1$ so that the AR(2) can be written

$$z_t - (1-\beta_1)z_{t-1} = \beta_3 + e_t$$

where $z_t = (1-B)y_t$. Thus the first difference follows an AR(1) model.

Box and Jenkins choose between competing models by sampling theoretic criteria such as goodness of fit tests and residual sum of squares. From the Bayesian viewpoint model selection can be accomplished by the computation of posterior probabilities. The forecast functions of the various models can be combined in the standard method. Namely, the predictive density of the future observations $\tilde{y}$ is given by

$$p(\tilde{y}|y) \propto \sum_s p(s|y)p(\tilde{y}|y,s)$$

where $p(\tilde{y}|y,s)$ is the predictive density of $\tilde{y}$ given constraints s are non-binding and $p(s|y)$ is the posterior probability of constraints s being non-binding.

This example is continued in section 7 where series C from [2] is studied.

3.2 Polynomial Regression

Stevens [14] derives the model

$$\Delta w/(1-x) = \sum_{j=1}^{p} \beta_j x^{j-1} \tag{3.1}$$

where Δw is the difference in weight concentrations of a solute and x is a dimensionless constant which can be considered known [14, equation 2]. The parameter space is (2.2) with

$C = (c_{ij})$ a $p \times p$ matrix with $c_{ij} = \begin{cases} 1 & i=j \\ -1 & j=i+1 \\ 0 & \text{otherwise} \end{cases}$

and $f=0$ [14, equation 4]. If the i^{th} constraint is binding $(\beta_i = \beta_{i+1})$, a polymer of i times the basic molecular weight is not present in the particular molecule.

Based on equation (3.1) either of the following models may be appropriate

(a) $$y_i = \Delta w_i/(1 - x_i) = \sum_{j=1}^{p} \beta_j x_i^{j-1} + e_i$$

(b) $$y_i = \Delta w_i = \sum_{j=1}^{p} \beta_j x_i^{j-1}(1 - x_i) + e_i$$

where $e \sim \mathcal{N}(0, \tau^{-1} I)$. These models are both of the form (2.1), and model (a) is the standard polynomial regression model with restricted parameter space.

The choice of the order p of the polynomial regression is also of interest in this problem. Halpern [6] discussed choice of the order in the unrestricted polynomial regression using the natural normal/gamma prior. The analogous method could be used in the restricted case.

3.3 Transition Probability Estimation

Judge and Takayama [10, pp. 176-8] use quadratic programming to obtain the least squares estimate of the transition probabilities of a finite Markov Chain. We consider the chain with 3 states and demonstrate how the theory of this paper applies. Judge and Takayama show the equations

$$y_{j,t} = \sum_{i=1}^{3} y_{i,t-1}\, p_{ij} + u_{jt} \quad \text{for} \quad i \leq j \leq 3,\ 1 \leq t \leq n$$

can be written in the form

$\tilde{y} = \tilde{X}\, p + u$ with

$\tilde{y}' = (y_1', y_2', y_3')$, $p' = (p_1', p_2', p_3')$, $u' = (u_1', u_2', u_3')$,

$y_j' = (y_{j,1}, \ldots, y_{j,n})$, $p_j' = (p_{1j}, p_{2j}, p_{3j})$, $u_j' = (u_{j1}, \ldots, u_{jn})$,

$$\tilde{X} = \begin{pmatrix} X_1 & 0 & 0 \\ 0 & X_1 & 0 \\ 0 & 0 & X_1 \end{pmatrix} \text{ where } X_1 = (z_1\ z_2\ z_3) \text{ and}$$

$z_j' = (y_{j,0}, \ldots, y_{j,n-1})$.

Since $p_1 + p_2 + p_3 = 1$, one can eliminate p_3 and write these equations as (2.1) with

$$y' = (y_1', y_2', y_3' - 1'X'), \quad X = \begin{pmatrix} X_1 & 0 \\ 0 & X_1 \\ -X_1 & X_1 \end{pmatrix} ,$$

$\beta' = (p_1', p_2')$, and $e = u$. The inequality restrictions are of the form (2.2) with

$$C = \begin{pmatrix} & I_6 & \\ -I_3 & & -I_3 \end{pmatrix} \quad \text{and} \quad f' = (0', -1_3')$$

where I_k is the $k \times k$ identity matrix and 1_k is a k vector of ones.

3.4 One Way Layout

Consider the 1 way fixed effect ANOVA model

$$y_{ij} = \beta_i + e_{ij} \qquad i=1,\ldots,p \quad \text{and} \quad j=1,\ldots,n_i$$

with β_i fixed and e_{ij} i.i.d. $\sim \mathcal{N}(0,\tau^{-1})$. We assume a priori that the means are non-decreasing (i.e., $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_p$). This constraint can be written in the form (2.2) with $C = (c_{ij})$ a $(p-1) \times p$ dimensional matrix with

$$c_{ij} = \begin{cases} 1 & j=i+1 \\ -1 & i=j \\ 0 & \text{Otherwise} \end{cases}$$

and $f=0$. An example of the use of this model is given in [1, example 1.3].

Other linear restrictions can be handled similarly. In the case of monotone decreasing parameters, Lindley [11] discussed the posterior distribution of the parameters assuming a vague prior.

4. The Likelihood

The following theorem expresses the likelihood in a form which suggests the natural conjugate prior of the mixed type to be employed. It shows that the likelihood for $\beta \in \bar{s}$ can be specified by $\alpha = p-r+k$ parameters which, in general, can be picked in different ways.

Theorem 1

For any set s of nonbinding constraints there exists a subset $\{j_1,\ldots,j_\alpha\}$ of $\{1,2,\ldots,p\}$ such that if $\eta_i = \beta_{j_i}$ and $\gamma' = (\eta_1,\ldots,\eta_\alpha)$ then the likelihood (2.3) can be expressed as

$$\ell(s,\gamma,\tau\,|\,y) \propto \tau^{n/2} \exp\{-\tau(z-R\gamma)'(z-R\gamma)/2\}\ I(\Omega). \tag{4.1}$$

The event Ω can be expressed as the set of linear inequalities

$$\Omega = (D\gamma_2 > w) \tag{4.2}$$

where D is a $k \times \alpha_2$ dimensional matrix, $\gamma' = (\gamma_1',\gamma_2')$, α_i = dimension γ_i for $i=1$ and 2, $\alpha_1 = p - p_0$, and $\alpha_2 = p_0 + k - r$. If either of the dimensions of D is 0, $I(\Omega) \equiv 1$. The "n dimensional vector z is a function of X and y.

To be precise the quantities z, R, D, w, and γ should be subscripted to denote which of the possible sets of size α has been picked. We assume throughout that a rule has been established for picking the variables so that the subscript can be surpressed without confusion.

Proof: Since the matrix C_0 was assumed of full rank, the number of binding constraints satisfies $0 \leq r-k \leq p_0$. It is convenient to separate the proof into the following three cases.

Case 1 r-k=0 (no binding constraints)

Since $r \geq 1$ and $p_0 \geq 1$ by assumption, one has $\min\{\alpha_2, k\} \geq 1$ in this case. The proof follows by defining $z=y$, $R=(X_1\ X_2)$, $\gamma_i = \beta_i$ for i=1 and 2, $D = C$, and $w = f$.

Case 2 $1 \leq r-k < p_0$

We can pick columns $E = \{j_1, \ldots, j_{r-k}\}$ of $C_{\bar{s}}$ defined in (2.4), so that the resulting matrix is nonsingular. Let $\bar{E}$ be the complement of E (in $\{1,2,\ldots,p_0\}$) and define matrices

$$C_{11}(C_{21}) \text{ as columns } E \text{ of } C_{\bar{s}}(C_s) \text{ and similarly}$$

$$C_{12}(C_{22}) \text{ as columns } \bar{E} \text{ of } C_{\bar{s}}(C_s)$$

where $C'_s = (d_{i_1} \ \ldots \ d_{i_k})$. If $k=0$, C_{21} and C_{22} are degenerate.

Defining vectors $\beta'_3 = (\beta_{j_1}, \ldots, \beta_{j_{r-k}})$, $\gamma'_2 = (\beta_{j_{r-k+1}}, \ldots, \beta_{j_{p_0}})$ and $f'_s = (f_{i_1}, \ldots, f_{i_k})$, for $k > 0$ the constraints can be written

$$C_{11}\beta_3 + C_{12}\gamma_2 = f_{\bar{s}}$$

$$C_{21}\beta_3 + C_{22}\gamma_2 > f_s .$$

Using the above equations to eliminate β_3 one obtains

$$\beta_3 = C_{11}^{-1}(f_{\bar{s}} - C_{12}\gamma_2).$$

The nonbinding constraints can be written as $D\gamma_2 > w$ where $D = C_{22} - C_{21}C_{11}^{-1}C_{12}$ and $w = f_s - C_{21}C_{11}^{-1}f_{\bar{s}}$.

Now define $X_3(X_4)$ as columns $E(\bar{E})$ of X so that

$$X\beta = \sum_{2}^{4} X_i\beta_i = X_2\beta_2 + X_3C_{11}^{-1}(f_{\bar{s}} - C_{12}\gamma_2) + X_4\gamma_2 = R\gamma + X_3C_{11}^{-1}f_{\bar{s}}$$

with $R = (X_2\ X_4 - X_3C_{11}^{-1}C_{12})$ and $\gamma' = (\beta_2', \gamma_2')$. Thus $y - X\beta = z - R\gamma$ with $z = y - X_3C_{11}^{-1}f_{\bar{s}}$. The rest of the proof follows easily in this case. For $k = 0$ the proof is analogous.

<u>Case 3</u> $r-k = p_0$ (restricted parameters completely specified)

In this case $\bar{E}$ is empty so that C_{12}, C_{22}, and γ_2 as defined in case 2 don't exist. One has that

$$\beta_3 = C_{11}^{-1}f_{\bar{s}}$$

and the inequality constraints can be written as

$$C_{21}C_{11}^{-1}\ f_{\bar{s}} > f_s$$

which determine whether the constraints can be binding together or not.

Letting $R = X_2$, $\gamma = \beta_2$, and $z = y - X_3C_{11}^{-1}f_{\bar{s}}$ the result follows.

5. The Prior

The likelihood of Theorem 1 suggests a prior of the mixed form. Let $p(s)$ denote the a priori probability of the set of constraints s being non-binding. Since $p(\gamma,\tau|s)$ determines $p(\beta,\tau|s)$, it suffices to specify $p(\gamma,\tau|s)$, which we assume to be the natural conjugate prior suggested by Theorem 1. Probabilities of events and distributions of parameters can be calculated in the usual way. The following theorem evaluates the constant of integration for the natural conjugate prior. This constant is necessary to calculate the Bayes factor for determining posterior odds of the different sets of binding constraints.

Theorem 2. Let $\gamma' = (\gamma_1', \gamma_2')$ where γ_i has dimension α_i and consider a prior of the following form

$$p(\gamma,\tau|s) = c\ \tau^{(\upsilon+\alpha)/2-1}\exp\{-\tau(b\upsilon + (\gamma-a)'V(\gamma-a))/2\}\ I(\Omega) \qquad (5.1)$$

where $V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$ is a positive definite, symmetric matrix,

V_{ij} has dimension $\alpha_i \times \alpha_j$ $1 \leq i,\ j \leq 2$, $a' = (a_1', a_2')$ where a_i has dimension α_i and

$$I(\Omega) = \begin{cases} 1 & \min\{\alpha_2, k\} = 0 \\ I(D\gamma_2 > w) & \text{otherwise} \end{cases} .$$

Then the constant of integration c is given by

$$c^{-1} = \begin{cases} c_1 & \min\{\alpha_2, k\} = 0 \\ c_1 P(\Omega) & \text{otherwise} \end{cases}$$

where

$$c_1 = \Gamma(\upsilon/2)(2\pi)^{\alpha/2}/|V|^{1/2}(\upsilon b/2)^{\upsilon/2}$$

and $P(\Omega)$, with Ω as in (4.2), is computed using the distribution $\gamma_2 \sim T_{\alpha_2}(a_2, V_{22.1}/b, \upsilon)$. That is, γ_2 is an α_2 dimensional multivariate t random variable with mean a_2, precision $V_{22.1}/b$, and υ degrees of freedom [4, pp. 59-61]. The dimension of the multivariate t random variable often is omitted in the sequel.

<u>Proof</u>: If $\min\{\alpha_2, k\} = 0$, the parameter space is unrestricted so that the prior is the standard normal/gamma and the constant is easily evaluated. For the case $\min\{\alpha_2, k\} > 0$, integrating the unrestricted parameters γ_1 one obtains

$$c^{-1} = (2\pi)^{\alpha_1/2} |V_{11}|^{-\frac{1}{2}} \iint \tau^{(\nu+\alpha_0)/2} \exp\{-\tau((\gamma_2-a_2)' V_{22.1}(\gamma_2-a_2) + b\nu)/2\}$$

$$I(\Omega) d\tau \, d\gamma_2$$

where $V_{22.1} = V_{22} - V_{21} V_{11}^{-1} V_{12}$. Integrating out τ and recognizing the multivariate t form, one has

$$c^{-1} = (2\pi)^{(\alpha_1+\alpha_2)/2} \Gamma(\nu/2)(|V_{11}||V_{22.1}|)^{-\frac{1}{2}}(\nu b/2)^{-\nu/2} \int_\Omega f(\gamma_2 | a_2, V_{22.1}/b, \nu) d\gamma_2$$

where $f(\gamma_2 | a_2, V_{22.1}/b, \nu)$ is the density of γ_2 where $\gamma_2 \sim T_{\alpha_2}(a_2, V_{22.1}/b, \nu)$. Thus $c^{-1} = c_1 P(\Omega)$ where $P(\Omega)$ is computed as in the statement of the Theorem.

The following corollary shows that $P(\Omega)$ of Theorem 2 can be computed (if $r \leq p_0$) using the c.d.f. of the multivariate t distribution with mean vector equal to 0 and diagonal elements of the precision matrix equal to 1. Approximations to the c.d.f. in this case have been given by John [7]. In the case $r > p_0$, $D\gamma_2$ has a degenerate multivariate t distribution so it is more convenient to compute $P(\Omega)$ using the α_2 dimensional distribution of γ_2 (rather than the distribution of $D\gamma_2$).

Corollary 2.1

If $r \leq p_0$, $P(\Omega) = F(-w|-Da_2, (DV_{22.1}^{-1}D')^{-1}/b, \nu)$ where $F(\cdot|a,P,\nu)$ is the c.d.f. of a $T(a,P,\nu)$ distribution. Furthermore, if the precision matrix $\Sigma = (DV_{22.1}^{-1}D')^{-1}/b = (\sigma_{ij})$ is written as $\Lambda\Delta\Lambda'$ with $\Lambda = (\sigma_{ii}^{\frac{1}{2}}\delta_{ij})$ where δ_{ij} is the Kronecker delta and

$$\Delta = (\sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{\frac{1}{2}}) \text{ then}$$

$$P(\Omega) = F(\Lambda(Da_2 - w)|0,\Delta,\nu).$$

Proof: If $k \leq \alpha_2 = p_0 - r + k$, then $D\gamma_2 \sim T_k(Da_2, (DV_{22.1}^{-1}D')^{-1}/b, \nu)$ (see, for example, [13, Theorem 6.2.1]). Thus, if $p_0 \geq r$, $P(\Omega) = F(-w|-Da_2, (DV_{22.1}^{-1}D')^{-1}/b, \nu)$. Since $\Lambda D(\gamma_2 - a_2) \sim T(0,\Delta,\nu)$, $P(\Omega) = F(\Lambda(Da_2 - w)|0,\Delta,\nu)$.

As stated above in the case $r > p_0$, it is more convenient to choose a different transformation from that given in Corollary 2.1. If the precision matrix $\Sigma = V_{22.1}/b$ is factored as in Corollary 2.1 $(=\Lambda\Delta\Lambda')$, then $u = \Lambda\ (\gamma_2 - a_2) \sim T(0,\Delta,\nu)$. Under this transformation the region $\Omega = (D\Lambda^{-1}u > w - Da_2)$. Thus, the inequalities on u are linear so that the region is a polygon or an infinite region with planer boundaries. John [8,9] has given methods for approximating the probability $P(u \in \Omega)$ where u has the above distribution.

In the case $\min\{\alpha_2, k\} = 0$ the prior of Theorem 2 reduces to the standard normal/gamma. In the restricted case the distribution of the parameters is more complicated and is given in the following corollary.

Corollary 2.2

If $\min\{\alpha_2, k\} > 0$, conditional distributions of the parameters γ_1 and γ_2 are

(i) $\gamma_2 | s \sim T_{\alpha_2}(a_2, V_{22.1}/b, \nu) I(\Omega)$, a multivariate t distribution, with the parameter as given, truncated to the space (4.2).

(ii) The p.d.f. $p(\gamma_1 | s)$ of γ_1 given s is
$p(\gamma_1 | s) = f(\gamma_1 | a_1, V_{11.2}/b, \nu)\, g(\gamma_1)$
where f is the density of the multivariate t and
$g(\gamma_1) = P(D\gamma_2 > w)$ where $\gamma_2 \sim T(a_{2.1}, (\nu+\alpha_1)V_{22}/(b\nu + (\gamma_1 - a_1)' V_{11.2}(\gamma_1 - a_1)), \nu + \alpha_1)$ and $a_{2.1} = a_2 - V_{22}^{-1} V_{21} (\gamma_1 - a_1)$.

Proof: (i) was shown in proving Theorem 2.

(ii) By integrating out τ one has

$$p(\gamma | s) \propto ((\gamma - a)' V (\gamma - a) + \nu b)^{-(\nu+\alpha)/2}\, I(\Omega).$$

Using the identity

$$(\gamma - a)' V (\gamma - a) = (\gamma_1 - a_1)' V_{11.2} (\gamma_1 - a_1) + (\gamma_2 - a_{2.1})' V_{22} (\gamma_2 - a_{2.1})$$

one has

$$p(\gamma_1 | s) \propto ((\gamma_1 - a_1)' V_{11.2} (\gamma_1 - a_1) + \nu b)^{-(\nu+\alpha_1)/2}$$

$$\int_{\Omega}(\nu b+(\gamma_1-a_1)'V_{11.2}(\gamma_1-a_1))^{-\alpha_2/2}\left(1+\frac{(\gamma_2-a_{2.1})'V_{22}(\gamma_2-a_{2.1})}{\nu b+(\gamma_1-a_1)'V_{11.2}(\gamma_1-a_1)}\right)^{-(\nu+\alpha)/2} d\gamma_2 \qquad (5.2)$$

The first term on the right hand side of (5.2) is proportional to $f(\gamma_1|a_1,V_{11.2}/b,\nu)$ and the second is proportional to $g(\gamma_1)$.

It is instructive to compare these distributions with the ones obtained in the unrestricted case (i.e., $I(\Omega)\equiv 1$ in (5.1)). In the unrestricted case the distribution of $\gamma_2|s$ is $T(a_2,V_{22.1}/b,\nu)$ which is the same as in (i) modulo the restriction. Similarly, the density of $\gamma_1|s$ in the restricted problem equals its density in the unrestricted problem times the function $g(\gamma_1)$.

Corollary 2.2 shows that picking the constants (a,V,ν,b) of the prior (5.1) is considerably more complicated than picking the constants in the unrestricted case. The following simpler prior is useful when prior knowledge about the parameters (given the constraints) is vague. Box and Tiao [2] use similar priors as reference priors. Of course, in repeated experiments the posterior of the first experiment can be used as a prior distribution for the second. Since the prior (5.1) is conjugate for the likelihood (4.1), it is useful in analysis of repeated experiments.

Corollary 2.3

Let Ω be as in Theorem 2 and assume that the Lebesque measure $m(\Omega)<\infty$. For a prior density of the form

$$p(\gamma,\tau|s)=c_0\tau^{(\nu+\alpha_1)/2-1}\exp\{-\tau(b\nu+(\gamma_1-a_1)'V_{11}(\gamma_1-a_1))/2\}\, I(\Omega) \qquad (5.3)$$

then one has

$$c_0^{-1}=\begin{cases} c_2 m(\Omega) & \min\{k,\alpha_2\}\geq 1 \\ c_2 & \min\{k,\alpha_2\}=0\end{cases}$$

with $c_2 = (2\pi)^{\alpha_1/2}\,\Gamma(\nu/2)/|V_{11}|^{1/2}\,(b\nu/2)^{\nu/2}$.

Proof: The integral of (5.3) is easily evaluated since (γ_1, τ), which has the standard normal/gamma form, is independent of γ_2.

Note that the constants of the prior (5.3), unlike (5.1) can easily be picked since

$$\tau | s \sim \Gamma_{\nu/2, b\nu/2}$$

and

$$\gamma_1 | s \sim T_{\alpha_1}(a_1, V_{11}/b, \nu).$$

The prior (5.3) can be obtained from the more general (5.1) by setting the elements of V_{12}, V_{22}, and a_2 equal to 0, changing α to α_1 and modifying the constant of integration. Similarly, the posterior distribution using prior (5.3) can be obtained from the posterior using prior (5.1) by the same specification. This will be utilized in the next section.

6. The Posterior

Since the conjugate prior distribution was assumed in section 5, the posterior distribution is of the same form as the prior. Thus, distribution results obtained for the prior distribution (i.e., Corollary 2.2) are valid for the posterior distribution with modified constants. Theorem 3 gives the rule for updating the constants in going from the prior to the posterior distribution.

Theorem 3. Using the likelihood (4.1) and the prior (5.1) one obtains a mixed posterior with

$$p(\gamma, \tau | y, s) \propto \tau^{(\overline{\nu}+\alpha)/2-1} \exp\{\tau((\gamma-\overline{a})'\overline{V}(\gamma-\overline{a}) + \overline{\nu}\overline{b})/2\}\, I(\Omega) \qquad (6.1)$$

with constants $\bar{\nu} = \nu + n$, $\bar{V} = V + R'R$, $\bar{a} = \bar{V}^{-1}(Va + R'z)$, $\bar{b} = (\nu b + Q)/\bar{\nu}$ where

$Q = (z - R\hat{\gamma})'(z - R\hat{\gamma}) + (a - \bar{a})'V(a - \bar{a}) + (\hat{\gamma} - \bar{a})'R'R(\hat{\gamma} - \bar{a})$ and $\hat{\gamma}$ is the OLS estimate $(R'R)^{-1}R'z$.

Proof: Follows by combining the likelihood with the prior and completing the square in the exponent (see [16, p.308] for a calculation of Q).

Using prior (5.3) the posterior density $p(\gamma, \tau \mid s, y)$ is of the form (6.1) but with the following modifications in the constants. Let $V = \begin{pmatrix} V_{11} & 0 \\ 0 & 0 \end{pmatrix}$ and $a' = (a_1', 0')$ then $\bar{V}, \bar{a}, \bar{b}$, and Q are as given in Theorem 3 and $\bar{\nu} = n + \nu - \alpha_2$.

The following theorem gives (up to a multiplicative constant) the posterior probability of the binding constraints. Zellner [15, section 10.4] gives a similar analysis for unconstrained linear regression.

Theorem 4. Using likelihood (4.1) and the prior (5.1) the posterior probability $p(s \mid y)$ of constraints s being non-binding is given by $p(s \mid y) \propto c_3(s)\, p(s)$ where

$$c_3 = c_3(s) = \begin{cases} c_4\, \bar{P}(\Omega)/P(\Omega) & \min\{\alpha_2, k\} \geq 1 \\ c_4 & \min\{\alpha_2, k\} = 0 \end{cases} .$$

$P(\Omega)$ is computed as in Theorem 2, $\bar{P}(\Omega)$ is the probability of the same event using the distribution $\gamma_2 \sim T(\bar{a}_2, \bar{V}_{22.1}/\bar{b}, \bar{\nu})$ where the constants are given in Theorem 3, and

$$c_4 = \Gamma(\bar{\nu}/2)(\nu b)^{\nu/2}|V|^{1/2} / \left(\Gamma(\nu/2)(\bar{\nu}\bar{b})^{\bar{\nu}/2}|\bar{V}|^{1/2}\right)$$

Proof: For $\min\{\alpha_2, k\} \geq 1$ one has

$$p(s,\gamma,\tau|y) \propto cp(s)\,\tau^{(\nu+\bar{\alpha})/2-1} \exp\{-\tau(\overline{b\nu}+(\gamma-\bar{a})'\bar{V}(\gamma-\bar{a}))/2\}I(\Omega)$$

where $c = c(s)$ is given in Theorem 2. Using the method of Theorem 2 to integrate out (γ,τ) one obtains

$$p(s|y) \propto p(s)\,\Gamma(\bar{\nu}/2)(2\pi)^{\alpha/2}\bar{P}(\Omega)/(|\bar{V}|^{1/2}(\bar{\nu}\bar{b}/2)^{\bar{\nu}/2} c_1 P(\Omega))$$

$$\propto c_4\bar{P}(\Omega)/P(\Omega).$$

A similar argument gives the result in the case $\min\{\alpha_2, k\} = 0$.

It follows easily from Theorem 4 that the Bayes factor of model s_i to model s_j is given by $c_3(s_i)/c_3(s_j)$. The rest of this section deals with the case when the prior information is vague. Corollary 4.1 specializes the result of Theorem 4 to the case when using prior (5.3).

Corollary 4.1. Using likelihood (4.1) and the prior (5.3),

$$p(s|y) \propto c_3(s)p(s)$$

with $c_3 = \begin{cases} c_4\,\bar{P}(\Omega)/m(\Omega) & \min\{k,\alpha_2\} \geq 1 \\ c_4 & \min\{k,\alpha_2\} = 0 \end{cases}$

with

$$c_4 = \Gamma(\bar{\nu}/2)|V_{11}|^{1/2}(\nu b)^{\nu/2}\,\pi^{\alpha_2/2}/(\Gamma(\nu/2)|\bar{V}|^{1/2}(\bar{\nu}\bar{b})^{\bar{\nu}/2})$$

with $\bar{\nu},\bar{V},\bar{b}$ as defined following Theorem 3 and $\bar{P}(\Omega)$ as in Theorem 4 with modified parameters.

Proof: For $\min\{\alpha_2,k\} \geq 1$ one has

$$p(\gamma,\tau,s|y) \propto c_0\tau^{(\bar{\nu}+\alpha)/2-1} \exp\{-\tau(\bar{b}\bar{\nu} + (\gamma-\bar{a})'\bar{V}(\gamma-\bar{a}))/2\}I(\Omega)$$

with $\bar{\nu},\bar{V},\bar{b}$ as above and c_0 given in Corollary 2.3. Integration with respect to (γ,τ) gives the result.

In many situations the prior opinion about (γ_1,τ) will not depend on s. In this case the expressions for the posterior probabilities of the various models are somewhat simpler.

Corollary 4.2. Using prior (5.3) and assuming b,ν, and V_{11} are the same for all s, then $p(s|y)$ is given as in Corollary 4.1 with

$$c_4 = \Gamma(\bar{\nu}/2)\ \pi^{\alpha_2/2}/(|\bar{V}|^{1/2}(\bar{\nu}\bar{b})^{\bar{\nu}/2})$$

Proof: The quantities $|V_{11}|^{1/2}$, $(\nu b)^{\nu/2}$, and $\Gamma(\nu/2)$ are the same.

Before experimentation, knowledge about γ_1 and τ may also be vague. This situation can be characterized by small values of b,ν, and V_{11}. We can approximate the situation by letting all these quantities converge to zero. By Scheffé's Theorem, the limiting values then serve as approximate probabilities in the case of vague prior information.

Corollary 4.3. The limit of c_4 (in corollary 4.2) as $\nu \to 0$, $b \to 0$, and $V_{11} \to 0$ (i.e., all coordinates go to 0) is given by

$$c_4 = \Gamma((n-\alpha_2)/2)\ \pi^{\alpha_2/2}/|R'R|^{1/2}\ Q^{(n-\alpha_2)/2}$$

where $Q = (z-R\hat{\gamma})'(z-R\hat{\gamma})$.

Furthermore, $\overline{P}(\Omega)$ of corollary 4.1 converges to $P(D\gamma_2 > w)$ where $\gamma_2 \sim T(\hat{\gamma}_2, (n-\alpha_2)\overline{V}_{22.1}/Q, n-\alpha_2)$ and $\overline{V}_{22.1}$ is computed from $\overline{V} = R'R$.

Proof: Taking the limit of the above expressions, one has $\overline{\nu} \to n - \alpha_2$, $\overline{V} \to R'R$, $\overline{\nu b} \to Q$, and $\overline{a} \to \hat{\gamma}$.

A reasonable sampling theoretic approach to model selection is through maximum likelihood. Although the m.l.e. of (β,τ) can be computed using quadratic programming, to contrast the Bayesian and sampling methods it is instructive to calculate the m.l.e. of (γ,τ,s). For fixed s the unrestricted maximum of the logarithm of the likelihood (4.1) is (ignoring an additive constant)

$$q(s) = -n(1 + \ln Q/n)/2$$

where $Q = (z-R\hat{\gamma})'(z-R\hat{\gamma})$ and $\hat{\gamma}$ is the OLS estimate. The m.l.e. of s maximizes $q(s)$ (i.e., minimizes the residual sum of squares) over those s such that $(D\hat{\gamma}_2 \geq w)$. If any of the theoretical constraints are violated using the OLS estimates for the parameters, the model is not feasible and can't be selected. This is somewhat unfortunate since sampling error could lead to violation of a constraint if the true parameter is near the boundary of the parameter space.

The Bayesian method does not completely rule out models which have a constraint violated by the OLS estimate. Gross violation of a constraint does lead to a small posterior probability. For example, using the approximate distribution of Corollary 4.3

$$P(\Omega) \approx F(\Lambda(D\hat{\gamma}_2 - w)|0,\Delta,n-\alpha_2)$$

where Λ and Δ are determined as an in Corollary 2.1 with precision matrix $(n-\alpha_2)\overline{V}_{22.1}/Q$ from Corollary 4.3. If a component of $\Lambda(D\hat{\gamma}_2-w)$ is less than -3, the posterior probability of that model will be small.

Corollary 4.3 shows that the Bayesian method of model selection also leads to picking models that have small residual sums of squares (i.e., small Q).

The next section applies this theory to a time series of temperatures of a chemical reaction [2, series C].

7. Analysis of Chemical Temperature Data

Computation of the posterior probabilities of binding constraints is given now for 226 successive readings of temperatures of a chemical process [2, series C]. For this series Box and Jenkins tentatively identify the number of differences to be $d = 1$ or 2 [2, p.185]. The model with $d = 2$ (even when overfit) is untenable based on the χ^2 goodness of fit test, but the model with $d = 1$ passes all the diagnostic tests [2, pp.292-3].

It is reasonable to restrict attention to AR(2) models only after looking at the partial autocorrelation function. Thus, the prior distribution is chosen in this case after some data analysis. Dickey [5] advocates using an approximate prior after some "data crunching."

After the identification stage of Box and Jenkins, one might have strong opinions about the parameters. As an operational (or reference) prior, we employ the vague prior (5.3). The constants $c_4(s)$, which determine the Bayes factors, are calculated from the limiting values of Corollary 4.3. Table 1 gives the value of $\ln(p(s|y)/p(s))$ for the seven possible models. The logarithm of the Bayes factor for model s_i to s_j is given by $\ln(p(s_i|y)/p(s_i)) - \ln(p(s_j|y)/p(s_j))$. Thus the Bayes factors can be computed by subtraction and then exponentiation. Using the Bayes factor one can easily calculate posterior probabilities for any choice of $p(s)$. The values for $p(s) = 1/7$ are given in Table 1.

-- Table 1 about here --

The table confirms the findings of Box and Jenkins that a single difference is best, but a probability of .3 is assigned to the model with no differences. The similarity in the estimates of the parameters (β_1, β_2) for the models corresponding to $d = 0$ and $d = 1$ shows why it is difficult to pick between these models. Of course, one might assign higher prior probability to higher differences (i.e., $p(\{1\}) > p(\text{none})$) since this is an uncontrolled reaction.

Table 1 Bayesian Analysis of Temperature of Chemical Reaction

$\bar{s}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	Q	$\ln(p(s\|y)/p(s))$	$p(s\|y)$
none	1.8017	-.8110	1.28×10^{-4}	3.8778	247.937	.300
1	1.8090	-.8090	5.53×10^{-5}	4.0026	248.752	.678
2	-.0069	.9931	-2.16×10^{-4}	42.959	-18.962	0
3	1.9903	-1	-1.79×10^{-4}	4.3059	237.863	0
1,2	0	1	-7.05×10^{-2}	43.005	-8.942	0
1,3	2	-1	-2.68×10^{-3}	4.4384	245.413	.022
2,3	-2	-1	11.91	14681	-662.237	0

Abstract

This paper considers the general linear model when the parameter space is subject to linear inequalities. Four examples satisfying the assumptions are given. A Bayesian analysis of this model is presented using a natural conjugate prior density of the mixed type. An analysis is given for the case that prior information about the parameters is vague. The Bayesian and sampling methods of model selection are compared in the case when little is known a priori.

References

[1] Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D., Statistical Inference Under Order Restrictions, New York: John Wiley and Sons, Inc., 1972.

[2] Box, G. E. P. and Jenkins, G. M., Time Series Analysis; Forecasting and Control, San Francisco: Holden-Day, Inc., 1970.

[3] Box, G. E. P. and Tiao, G. C., Bayesian Inference in Statistical Analysis, Reading,Mass.: Addison-Wesley Publishing Co., 1973.

[4] De Groot, M. H., Optimal Statistical Decisions, New York: McGraw-Hill Book Co., 1970.

[5] Dickey, J. M., "Approximation Posterior Distributions," to appear in J. Amer. Statist. Assoc., 71 (Sept. 1976).

[6] Halpern, E. F., "Polynomial Regression from a Bayesian Approach," J. Amer. Statist. Assoc., 68, (March 1973), 137-143.

[7] John, S., "On the Evaluation of the Probability Integral of the Multivariate t-Distribution," Biometrika, 48, No. 3 (1961), 409-417.

[8] John, S., "Methods for Evaluation of Probabilities of Polygonal and Angular Regions When the Distribution is Bivariate t," Sankhya A, 26, (July 1964), 47-54.

[9] John, S., "On the Evaluation of Probabilities of Convex Polyhedra Under Multivariate Normal and t Distributions," J. Roy. Statist. Soc. B, 28, (1966), 366-369.

[10] Judge, C. G. and Takayama, T., "Inequality Restrictions in Regression Analysis," J. Amer. Statist. Assoc., 61, (March 1966), 166-181.

[11] Lindley, D. V., Discussion of Bartholomew, D. J., "A Test of Homogeneity of Means Under Restricted Alternatives," J. Royal Statist. Soc. B, No. 2 (1961), 272-274.

[12] Lovell, M. C. and Prescott, E., "Multiple Regression with Inequality Constraints: Pretesting Bias, Hypothesis Testing and Efficiency," J. Amer. Statist. Assoc., 65, (June 1970), 913-925.

[13] Press, S. J., Applied Multivariate Analysis, New York: Holt, Rinehart and Winston, Inc., 1972.

[14] Stevens, C. L., "Equilibrium Ultracentrifugation of Polymers: A New Method of Data Reduction with an Application to TMV Protein," FEBS Letters, 56 Number 1, (August 1975), 12-15.

[15] Zellner, A., An Introduction to Bayesian Inference in Econometrics, New York: John Wiley and Sons, Inc., 1971.

[16] Zellner, A., "Linear Regression with Inequality Constraints on the Coefficients," International Center for Management Science Report Number 6109, 1961.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
ONR Report No. 4

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
6
Bayesian Analysis of the Linear Model Subject to Linear Inequality Constraints

5. TYPE OF REPORT & PERIOD COVERED
9
Technical Rep[redacted]t,

6. PERFORMING ORG. REPORT NUMBER
Technical Report No. 118

7. AUTHOR(s)
10
William W. Davis

8. CONTRACT OR GRANT NUMBER(s)
15
N00014-76-C-0930

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Department of Statistics
Carnegie-Mellon University
Pittsburgh, Pa. 15213

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
16
NR-042-367

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
Statistics and Probability Program
Code (436) Arlington, Va. 22217

12. REPORT DATE
11
Aug[redacted]76

13. NUMBER OF PAGES
27 12 30p.

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
14 TR-118, 4-ONR

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
linear model, linear inequality constraint, binding constraint, conjugate prior distrubution, vague prior, posterior probability, Bayes factor

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(see reverse side)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

391190 1B

This paper considers the general linear model when the parameter space is subject to linear inequalities. Four examples satisfying the assumptions are given. A Bayesian analysis of this model is presented using a natural conjugate prior density of the mixed type. An analysis is given for the case that prior information about the parameters is vague. The Bayesian and sampling methods of model selection are compared in the case when little is known a priori.